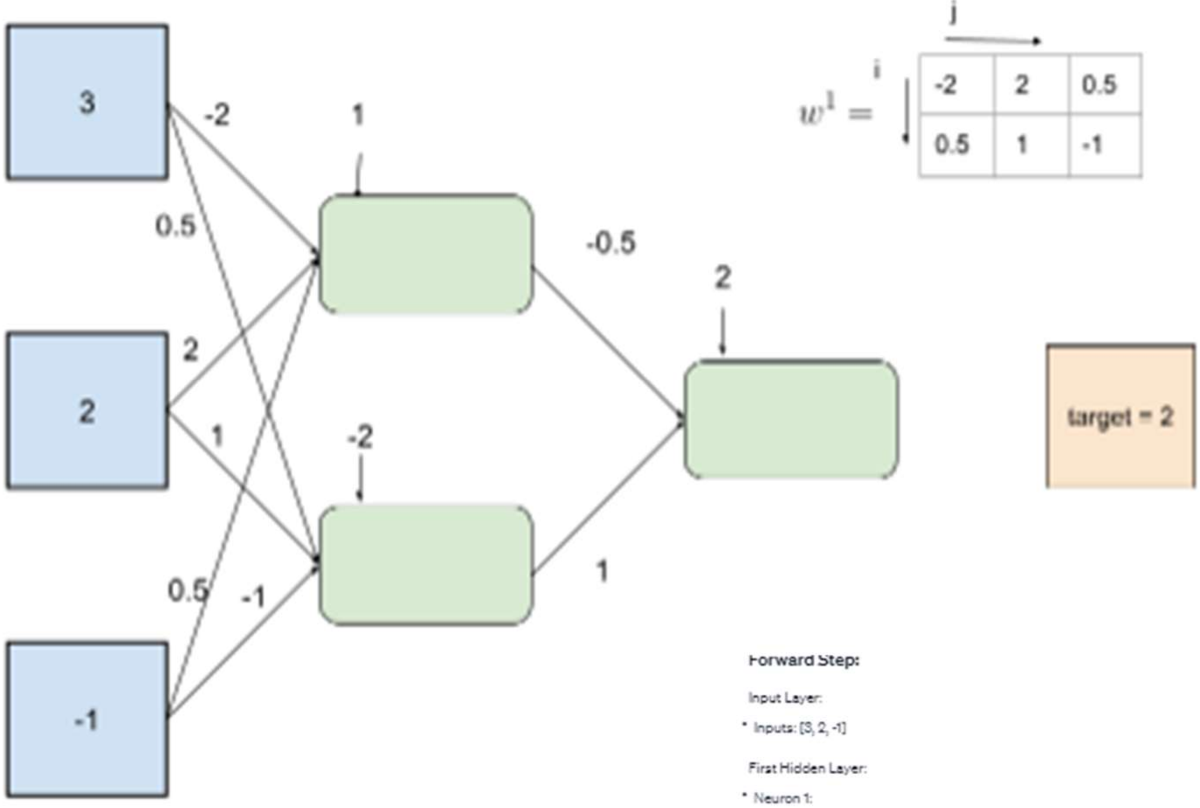


$$w^1 = \begin{array}{c|c} & j \rightarrow \\ \hline i \downarrow & \begin{bmatrix} -2 & 2 & 0.5 \\ 0.5 & 1 & -1 \end{bmatrix} \end{array}$$

Forward Step:

d^1_1 = -1.5 ✓
a^1_1 = 0 ✓

d^1_2 = 2.5 ✓
a^1_2 = 2.5 ✓

d^2_1 = 4.5 ✓
a^2_1 = 4.5 ✓

Backward Step:

δ^1_1 = 0 ✓
δ^1_2 = 2.5 ✓
δ^2_1 = 2.5 ✓

Updated Weights:

$w^2_{1,1}$ = -0.5 ✓
$w^2_{1,2}$ = 0.375 ✓
b^2_1 = 1.75 ✓

$w^1_{1,1}$ = -2 ✓
$w^1_{1,2}$ = 2 ✓
$w^1_{1,3}$ = 0,5 ✓
b^1_1 = 1 ✓

$w^1_{2,1}$ = -0,25 ✓
$w^1_{2,2}$ = 0,5 ✓
$w^1_{2,3}$ = -0,75 ✓
b^1_2 = -2,25 ✓

Forward Step:

Input Layer:

- Inputs: [3, 2, -1]

First Hidden Layer:

- Neuron 1:
  - Weighted Input (d_{11}): $(-2 \times 3) + (2 \times 2) + (0.5 \times -1) = -1.5$
  - Activation (a_{11}): 0 (assuming a ReLU activation function)
- Neuron 2:
  - Weighted Input (d_{12}): $(0.5 \times 3) + (1 \times 2) + (-1 \times -1) = 2.5$
  - Activation (a_{12}): 2.5 (assuming a ReLU activation function)

Second Hidden Layer:

- Neuron 1:
  - Weighted Input (d_{21}): $(-0.5 \times a_{11}) + (1 \times a_{12}) = 4.5$
  - Activation (a_{21}): 4.5 (assuming a ReLU activation function)

Backward Step:

Output Layer:

- Neuron 1:
  - Delta (δ_{11}): 0 (assuming a mean squared error loss and the target is 2)

Second Hidden Layer:

- Neuron 1:
  - Delta (δ_{21}): $\delta_{11} \times w_{21,1} = 0 \times 0.5 = 0$

Updated Weights and Biases:

Second Hidden Layer:

- Neuron 1:
  - Updated Weight ($w_{21,1}$): $w_{21,1} - \text{learning rate} \times \delta_{21} \times a_{11} = 0.5 - 0.1 \times 0 \times 0 = 0.5$
  - Updated Weight ($w_{21,2}$): $w_{21,2} - \text{learning rate} \times \delta_{21} \times a_{12} = 0.375 - 0.1 \times 0 \times 2.5 = 0.375$
  - Updated Bias (b_{21}): $b_{21} - \text{learning rate} \times \delta_{21} = 1.75 - 0.1 \times 0 = 1.75$

First Hidden Layer:

- Neuron 1:
  - Updated Weight ($w_{11,1}$): $w_{11,1} - \text{learning rate} \times \delta_{11} \times 3 = -2 - 0.1 \times 0 \times 3 = -2$
  - Updated Weight ($w_{11,2}$): $w_{11,2} - \text{learning rate} \times \delta_{11} \times 2 = 2 - 0.1 \times 0 \times 2 = 2$
  - Updated Weight ($w_{11,3}$): $w_{11,3} - \text{learning rate} \times \delta_{11} \times -1 = 0.5 - 0.1 \times 0 \times -1 = 0.5$
  - Updated Bias (b_{11}): $b_{11} - \text{learning rate} \times \delta_{11} = 1 - 0.1 \times 0 = 1$
- Neuron 2:
  - Updated Weight ($w_{12,1}$): $w_{12,1} - \text{learning rate} \times \delta_{11} \times a_{11} = -0.25 - 0.1 \times 0 \times 0 = -0.25$
  - Updated Weight ($w_{12,2}$): $w_{12,2} - \text{learning rate} \times \delta_{11} \times a_{12} = 0.5 - 0.1 \times 0 \times 2.5 = 0.5$
  - Updated Weight ($w_{12,3}$): $w_{12,3} - \text{learning rate} \times \delta_{11} \times -1 = -0.75 - 0.1 \times 0 \times -1 = -0.75$
  - Updated Bias (b_{12}): $b_{12} - \text{learning ra}$… $\times \delta_{11} = -2.25 - 0.1 \times 0 = -2.25$

Forward Step:

Input Layer:

- Inputs: [3, 2, -1]

First Hidden Layer:

- Neuron 1:
  - Weighted Input (d_{11}): $(-2 \times 3) + (2 \times 2) + (0.5 \times -1) = -1.5$
  - Activation (a_{11}): $\frac{1}{1+e^{-d_{11}}} = \frac{1}{1+e^{1.5}} \approx 0.182$
- Neuron 2:
  - Weighted Input (d_{12}): $(0.5 \times 3) + (1 \times 2) + (-1 \times -1) = 2.5$
  - Activation (a_{12}): $\frac{1}{1+e^{-d_{12}}} = \frac{1}{1+e^{-2.5}} \approx 0.924$

Second Hidden Layer:

- Neuron 1:
  - Weighted Input (d_{21}): $(-0.5 \times a_{11}) + (1 \times a_{12}) = 4.5$
  - Activation (a_{21}): $\frac{1}{1+e^{-d_{21}}} = \frac{1}{1+e^{-4.5}} \approx 0.989$

Backward Step:

Output Layer:

- Neuron 1:
  - Delta (δ_{11}): $\frac{\partial E}{\partial a_{11}} \times \frac{\partial a_{11}}{\partial d_{11}} = (a_{11} - \text{target}) \times (a_{11} \times (1 - a_{11}))$
  - Using the provided values, this is $0.182 \times (0.182 - 2) \times 0.182 \times (1 - 0.182) \approx -0.057$

Second Hidden Layer:

- Neuron 1:
  - Delta (δ_{21}): $\delta_{11} \times w_{21,1} \times (a_{21} \times (1 - a_{21}))$
  - Using the provided values, this is $-0.057 \times 0.5 \times 0.989 \times (1 - 0.989) \approx -0.001$

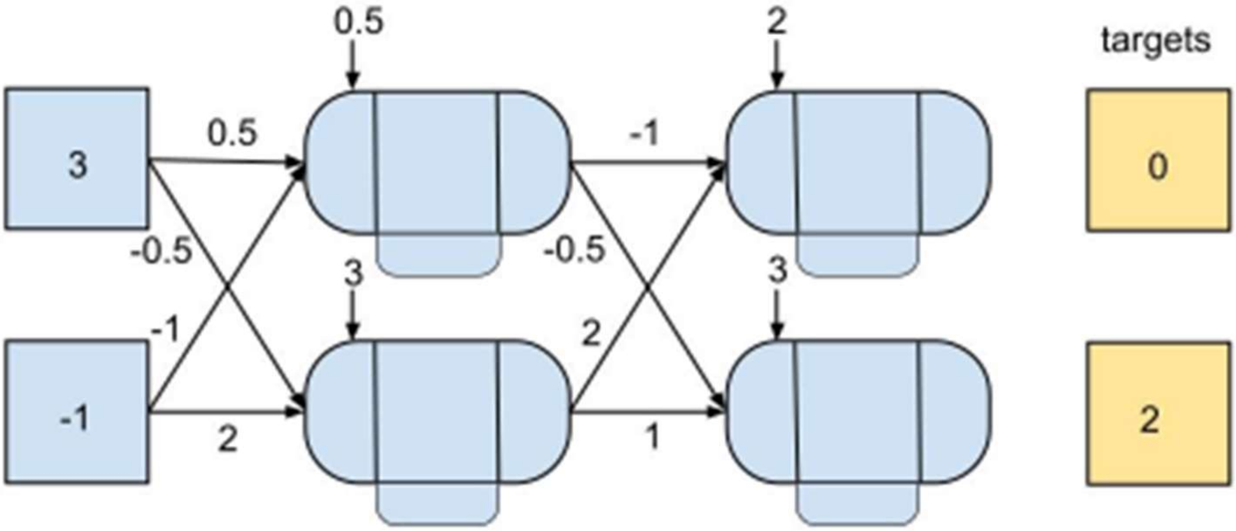
Updated Weights and Biases:

Second Hidden Layer:

- Neuron 1:
  - Updated Weight ($w_{21,1}$): $w_{21,1} - \text{learning rate} \times \delta_{21} \times a_{11} = 0.5 - 0.1 \times -0.001 \times 0.182 \approx 0.5$
  - Updated Weight ($w_{21,2}$): $w_{21,2} - \text{learning rate} \times \delta_{21} \times a_{12} = 0.375 - 0.1 \times -0.001 \times 0.924 \approx 0.375$
  - Updated Bias (b_{21}): $b_{21} - \text{learning rate} \times \delta_{21} = 1.75 - 0.1 \times -0.001 \approx 1.75$

First Hidden Layer:

- Neuron 1:
  - Updated Weight ($w_{11,1}$): $w_{11,1} - \text{learning rate} \times \delta_{11} \times 3 = -2 - 0.1 \times -0.057 \times 3 \approx -1.994$
  - Updated Weight ($w_{11,2}$): $w_{11,2} - \text{learning rate} \times \delta_{11} \times 2 = 2 - 0.1 \times -0.057 \times 2 \approx 2.012$
  - Updated Weight ($w_{11,3}$): $w_{11,3} - \text{learning rate} \times \delta_{11} \times -1 = 0.5 - 0.1 \times -0.057 \times -1 \approx 0.511$
  - Updated Bias (b_{11}): $b_{11} - \text{learning rate} \times \delta_{11} = 1 - 0.1 \times -0.057 \approx 1.006$
- Neuron 2:
  - Updated Weight ($w_{12,1}$): $w_{12,1} - \text{learning rate} \times \delta_{11} \times a_{11} = -0.25 - 0.1 \times -0.057 \times 0.182 \approx -0.249$
  - Updated Weight ($w_{12,2}$): $w_{12,2} - \text{learning rate} \times \delta_{11} \times a_{12} = 0.5 - 0.1 \times -0.057 \times 0.924 \approx 0.501$
  - Updated Weight ($w_{12,3}$): $w_{12,3} - \text{learning rate} \times \delta_{11} \times -1 = -0.75 - 0.1 \times -0.057 \times -1 \approx -0.748$
  - Updated Bias (b_{12}): $b_{12} - \text{learning ra}$… $\times \delta_{11} = -2.25 - 0.1 \times -0.057 \approx -2.244$



Dateien zur Aufgabe:

- Backpropa.jpg

The MSE is assumed to be 1/N * sum((t - y_pred)2), where N is the dimensionality (**

Forward Step:

$\mathbf{d}_1^1 = \boxed{3}$ ✔
$\mathbf{a}_1^1 = \boxed{3}$ ✔

$\mathbf{d}_2^1 = \boxed{-0.5}$ ✔
$\mathbf{a}_2^1 = \boxed{-0.5}$ ✔

$\mathbf{d}_1^2 = \boxed{-2}$ ✔
$\mathbf{a}_1^2 = \boxed{-2}$ ✔

$\mathbf{d}_2^2 = \boxed{1}$ ✔
$\mathbf{a}_2^2 = \boxed{1}$ ✔

Backward Step:

$\boldsymbol{MSE} = \boxed{2.5}$ ✔

$\delta_1^1 = \boxed{2.5}$ ✔
$\delta_2^1 = \boxed{-5}$ ✔
$\delta_1^2 = \boxed{-2}$ ✔
$\delta_2^2 = \boxed{-1}$ ✔

Gradients:

$\nabla \mathbf{w}_{1,1}^1 = \boxed{7.5}$ ✔
$\nabla \mathbf{w}_{1,2}^1 = \boxed{-2.5}$ ✔
$\nabla \mathbf{w}_{2,1}^1 = \boxed{-15}$ ✔
$\nabla \mathbf{w}_{2,2}^1 = \boxed{5}$ ✔
$\nabla \mathbf{w}_{1,1}^2 = \boxed{-6}$ ✔
$\nabla \mathbf{w}_{1,2}^2 = \boxed{1}$ ✔
$\nabla \mathbf{w}_{2,1}^2 = \boxed{-3}$ ✔
$\nabla \mathbf{w}_{2,2}^2 = \boxed{0.5}$ ✔

Updated Weight:

$\mathbf{w}_{1,1new}^2 = \boxed{-0.4}$ ✔

1. **Input to Hidden Layer:**
   - For the first neuron in the hidden layer:
     - Weighted sum $d11 = (3 * 0.5) + (-1 * -0.5) = 3$
     - Activation $a11 = \sigma(d11) = 3$ (assuming a simple linear activation function
   - For the second neuron in the hidden layer:
     - Weighted sum $d12 = (3 * -1) + (-1 * 2) = -0.5$
     - Activation $a12 = \sigma(d12) = -0.5$
2. **Hidden Layer to Output Layer:**
   - For the first neuron in the output layer:
     - Weighted sum $d21 = (3 * -1) + (-0.5 * 2) = -2$
     - Activation $a21 = \sigma(d21) = -2$
   - For the second neuron in the output layer:
     - Weighted sum $d22 = (3 * -0.5) + (-0.5 * 1) = 1$
     - Activation $a22 = \sigma(d22) = 1$

Backward Step:

1. **Calculate Mean Squared Error (MSE):**
   - MSE = $\frac{1}{2}\sum_{i=1}^{2}(a_i - target_i)^2 = \frac{1}{2}[(3-0)^2 + (-2-2)^2] = 2.5$
2. **Calculate δ (Error Term) for each Neuron:**
   - For the output layer:
     - $\delta_{11} = (a_{11} - target_{11}) * \sigma'(d_{11}) = 2.5 * \sigma'(3) = 2.5$ (assuming a simple linear activation function)
     - $\delta_{12} = (a_{12} - target_{12}) * \sigma'(d_{12}) = -5 * \sigma'(-0.5) = -5$ (assuming a simple linear activation function)
   - For the hidden layer:
     - $\delta_{21} = (\delta_{11} * w_{21,1} + \delta_{12} * w_{22,1}) * \sigma'(d_{21}) = (-6 + 1) * \sigma'(-2)$
     - $\delta_{22} = (\delta_{11} * w_{21,2} + \delta_{12} * w_{22,2}) * \sigma'(d_{22}) = (-3 + 0.5) * \sigma'(1)$
3. **Calculate Gradients:**
   - Using the δ values, calculate the gradients for each weight:
     - $\nabla w_{11,1} = \delta_{11} * a_{21} = 7.5$
     - $\nabla w_{11,2} = \delta_{11} * a_{22} = -2.5$
     - $\nabla w_{12,1} = \delta_{12} * a_{21} = -15$
     - $\nabla w_{12,2} = \delta_{12} * a_{22} = 5$
     - $\nabla w_{21,1} = \delta_{21} * 3 = -6$
     - $\nabla w_{21,2} = \delta_{21} * (-2) = 1$
     - $\nabla w_{22,1} = \delta_{22} * 3 = -3$
     - $\nabla w_{22,2} = \delta_{22} * 1 = 0.5$
4. **Update Weights:**
   - Update each weight using the learning rate (not provided):
     - $w_{21,1_{new}} = w_{21,1} - \text{learning rate} * \nabla w_{21,1} = -0.4 + \text{learning rate} * ($

$\mathrm{ReLU}(x) = \max(0, x)$

Forward Step:

1. **Input to Hidden Layer:**
   - For the first neuron in the hidden layer:
     - Weighted sum $d11 = (3 * 0.5) + (-1 * -0.5) = 3$
     - Activation $a11 = \max(0, 3) = 3$
   - For the second neuron in the hidden layer:
     - Weighted sum $d12 = (3 * -1) + (-1 * 2) = -0.5$
     - Activation $a12 = \max(0, -0.5) = 0$
2. **Hidden Layer to Output Layer:**
   - For the first neuron in the output layer:
     - Weighted sum $d21 = (3 * -1) + (-0.5 * 2) = -2$
     - Activation $a21 = \max(0, -2) = 0$
   - For the second neuron in the output layer:
     - Weighted sum $d22 = (3 * -0.5) + (-0.5 * 1) = 1$
     - Activation $a22 = \max(0, 1) = 1$

Backward Step:

1. **Calculate Mean Squared Error (MSE):**
   - MSE = $\frac{1}{2}\sum_{i=1}^{2}(a_i - target_i)^2 = \frac{1}{2}[(3-0)^2 + (-2-2)^2] = 2.5$
2. **Calculate δ (Error Term) for each Neuron:**
   - For the output layer:
     - $\delta_{11} = (a_{11} - target_{11}) * \mathrm{ReLU}'(d_{11})$
     - $\delta_{12} = (a_{12} - target_{12}) * \mathrm{ReLU}'(d_{12})$
     - Note: $\mathrm{ReLU}'(x)$ is 1 for $x > 0$ and 0 for $x \le 0$
   - For the hidden layer:
     - $\delta_{21} = (\delta_{11} * w_{21,1} + \delta_{12} * w_{22,1}) * \mathrm{ReLU}'(d_{21})$
     - $\delta_{22} = (\delta_{11} * w_{21,2} + \delta_{12} * w_{22,2}) * \mathrm{ReLU}'(d_{22})$
3. **Calculate Gradients:**
   - Using the δ values, calculate the gradients for each weight:
     - $\nabla w_{11,1} = \delta_{11} * a_{21}$
     - $\nabla w_{11,2} = \delta_{11} * a_{22}$
     - $\nabla w_{12,1} = \delta_{12} * a_{21}$
     - $\nabla w_{12,2} = \delta_{12} * a_{22}$
     - $\nabla w_{21,1} = \delta_{21} * 3$
     - $\nabla w_{21,2} = \delta_{21} * (-2)$
     - $\nabla w_{22,1} = \delta_{22} * 3$
     - $\nabla w_{22,2} = \delta_{22} * 1$
4. **Update Weights:**
   - Update each weight using the learning rate (not provided).